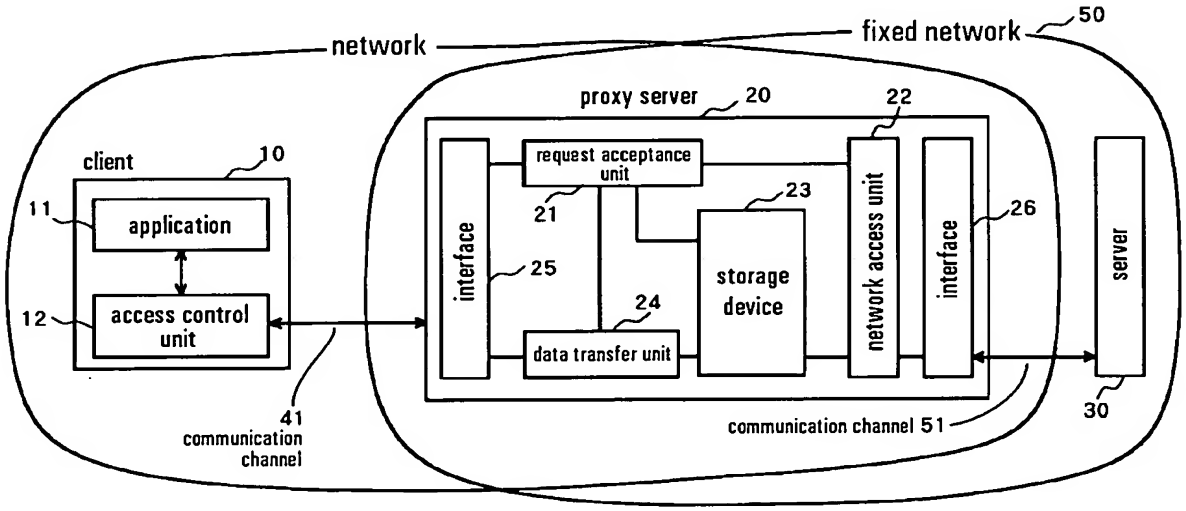
Translation of Japanese Unexamined Pat. Appl. Publication No. 10-312351

Application No.	10-054759
Filing Date	6 March 1998
Publication No.	10-312351
Publication Date	24 November 1998
Priority Claimed	Jap. Pat. Appl. No. 09-054813, 10 March 1997, Japan
Int. Cl.[6]	G06F 13/00; G06F 12/00; H04L 12/54; H04L 12/58
Inventors	Sadaichi IRIMIYA, Ushio INOUE and Satoshi HAKOMORI
Applicant	NTT Data Corporation

Title Data Communication System and Method, and Proxy Server for the System

Abstract



TASK: To shorten the duration of connection and thereby reduce communication cost when data is transferred through a network from a server to a client.

SOLUTION: Proxy server 20 is provided between client 10 and server 30. Proxy server 20 has a function whereby it interprets a data acquisition request from client 10, selects an appropriate server 30 to fulfil this request, and acquires and stores the requested data from this server 30. Proxy server 20 also has a function whereby in response to a data transfer request from client 10 it transfers, to client 10, requested data that has been stored. Client 10 connects to proxy server 20 through communication channel 41 only when sending a data acquisition request to proxy server 20, or when sending a data transfer request to proxy server 20 and receiving the requested data from proxy server 20. Communication channel 41 between client 10 and proxy server 20 is disconnected during the time when server 30 is preparing the requested data and while it is sending the prepared data to proxy server 20. Communication channel 51 between proxy server 20 and server 30 is preferably a stable, low-cost, broadband channel.

Claims

1. A data communication system comprising:

a client for issuing a data acquisition request;

one or more servers for supplying data; and

a proxy server for receiving the data acquisition request from said client, interpreting this data acquisition request and selecting a method suitable for acquisition of the requested data, and acquiring the requested data by means of the selected method and transferring it to said client.

2. The data communication system set forth in Claim 1, wherein the proxy server selects, as said method suitable for acquisition of the requested data:

A) if the data acquisition request has designated a server, a method wherein the proxy server acquires the requested data from the designated server; or

B) if the data acquisition request has not designated a server, a method wherein the proxy server selects, in accordance with a prescribed criterion, a server suitable for supplying the requested data, and acquires the requested data from the selected server.

3. The data communication system set forth in Claim 1 or Claim 2, wherein the proxy server selects, as said method suitable for acquisition of the requested data:

C) a method wherein, even if the data acquisition request has designated a server, if there is a more suitable server for providing the requested data than the designated server, the proxy sever acquires the requested data from this more suitable server.

4. The data communication system set forth in Claim 1, wherein the proxy server is provided with a storage device for storing requested data that has already been acquired, and wherein the proxy server selects, as said method suitable for acquisition of the requested data:

D) if the requested data is not present in the storage device when a data acquisition request is received, a method wherein the proxy server acquires the requested data from a server and transfers it to the client; or

E) if the requested data is already present in the storage device when a data acquisition request is received, a method wherein the proxy server omits the step of acquiring the requested data from a server and transfers, to the client, the requested data which is already present in the storage device.

5. The data communication system set forth in Claim 1, wherein the communication channel between the client and the proxy server:

F) is connected when the data acquisition request is sent from the client to the proxy server;

G) is disconnected after the transmission of the data acquisition request is finished; and

H) is reconnected when the requested data is transferred from the proxy server to the client.

6. The data communication system set forth in Claim 1 or Claim 5, wherein the communication channel between the proxy server and the server is always connected.

7. The data communication system set forth in Claim 1 or Claim 5, wherein the proxy server is provided with a storage device for storing requested data that has been acquired from a server, and transfers, to the client, the requested data in said storage device after all the requested data has been stored in the storage device.

8. The data communication system set forth in Claim 1, wherein the proxy server compresses and transfers the requested data.

9. A data communication method for transferring data from a server to the client in a data communication system comprising a client, one or more servers and a proxy server interposed between the client and the one or more servers, said data communication method comprising:

a step wherein said client issues a data acquisition request to said proxy server;

a step wherein said proxy server receives the data acquisition request, interprets this data acquisition request and selects a method suitable for acquisition of the requested data, and acquires the requested data by means of the selected method; and

a step wherein the proxy server transfers the acquired requested data to the client.

10. The data communication method set forth in Claim 9, wherein there is additionally provided a step wherein the communication channel between the client and the proxy server:

A) is connected when the data acquisition request is sent from the client to the proxy server;

B) is disconnected after the transmission of the data acquisition request is finished; and

C) is reconnected when the requested data is transferred from the proxy server to the client.

11. A proxy server for data communication, said proxy server being interposed between a client capable of issuing a data acquisition request and one or more servers for supplying data, and comprising:

a request acceptance means for receiving the data acquisition request from said client, and for interpreting this data acquisition request and selecting a method suitable for acquisition of the requested data;

an access means for acquiring the requested data by means of the method selected by said request acceptance means; and

a data transfer means for transferring, to said client, the requested data acquired by said access means.

12. The proxy server set forth in Claim 11, further provided with a storage means for holding said acquired requested data.

13. A computer-readable recording medium holding a program for causing a computer to function as a proxy server for data communication, said proxy server being interposed between a client capable of issuing a data acquisition request and one or more servers for supplying data, and comprising:

a request acceptance means for receiving the data acquisition request from said client, and for interpreting this data acquisition request and selecting a method suitable for acquisition of the requested data;

an access means for acquiring the requested data by means of the method selected by said request acceptance means; and

a data transfer means for transferring, to said client, the requested data acquired by said access means.

14. The recording medium set forth in Claim 13, further provided with a storage means for holding the requested data which said proxy server has acquired.

Detailed Description of the Invention

Technical field of the invention

(1) The present invention relates to data communication systems for transferring data from a server to a client in accordance with a request from the client. The invention can be advantageously utilised in online data acquisition from databases and in Internet-based information distribution services.*

Prior art

(2) FIG. 1 shows an example of the typical constitution of a conventional data communication system.

(3) Client 1 wishing to utilise data, and server 7 for supplying this data, are present on communication network 9. Client 1 and server 7 are connected whenever occasion demands via radio or wired communication channel 8. When application 3 in client 1 needs to acquire data from server 7, a procedure of the sort depicted in FIG. 2 is performed.

(4) Firstly, in client 1, a data acquisition request issued by application 3 is passed on to access control unit 5 (step S1). Access control unit 5 sends a connection request via communication channel 8 to server 7, and communication between the two is established by server 7 responding to this request (S2). After this, the data acquisition request issued by application 3 is conveyed from access control unit 5 via communication channel 8 to server 7 (S3).

(5) Server 7, responding to the data acquisition request, prepares the requested data and transfers this data via communication channel 8 to access control unit 5 of client 1 (S4). Access control unit 5 receives and stores the transferred data, and finally, when all the requested data has been received, informs application 3 of the result (S5).

* Numbers in round brackets at the beginning of paragraphs correspond to the paragraph numbering in the Japanese patent document.

Problem which the invention will solve

(6) In the system described above, communication channel 8 has to be kept connected throughout the period of time starting from a communication connection request being sent from client 1 to server 7, continuing while server 7 prepares data for transfer to client 1, and extending until all the requested data have finished being transferred to access control unit 5. Moreover, if the communication has been disconnected before the data transfer is completed, it is necessary for the procedure depicted in FIG. 2 to be carried out again, starting at least from step S2 where access control unit 5 sends a connection request to server 7. As a result, the total duration of connection between the client and the server becomes rather long.

(7) This may not be a major problem when the communication channel between the client and the server is a broadband channel with a low communication cost. However, the problem of communication cost mounting up is encountered when data communication is carried out using a communication channel with narrow bandwidth, high communication cost and unstable connections, as in the case for example of a wireless network or a public switched network. In particular, if there is heavy traffic or if the state of the channel deteriorates somewhere along the communication path between the client and the server, the time required for data acquisition and the trouble and cost involved will increase still more, due to reasons such as the data transfer rate decreasing and the frequent occurrence of disconnections.

(8) It is therefore an object of the present invention to provide a data communication system capable of transferring data from a server to a client by means of a shorter duration connection and hence at a lower communication cost, even when utilising a communication channel with narrow bandwidth, high communication cost and unstable connections.

Means for solving problem

(9) According to the present invention, a proxy server responsive to a data acquisition request issued by the client is provided between the client and a server. The proxy server interprets the data acquisition request from the client, selects a method suitable for acquiring the requested data, acquires the requested data by means of the selected method, and transfers it to the client.

(10) The methods which the proxy server selects are as follows:

(11) A) If the data acquisition request has designated a server, the proxy server acquires the requested data from the designated server.

(12) B) If the data acquisition request has not designated a server, the proxy server selects, in accordance with a prescribed criterion, a server suitable for supplying the requested data, and acquires the requested data from the selected server.

(13) C) Even if the data acquisition request has designated a server, if there is a more suitable server for providing the requested data than the designated server, the proxy server acquires the requested data from this more suitable server.

(14) It is also feasible to provide, in the proxy server, a storage device for holding requested data that has already been acquired, and for the proxy server to select a method of the following sort, these methods utilising this storage device:

(15) D) If the requested data is not present in the storage device when a data acquisition request is received, the proxy server acquires the requested data from a server and transfers it to the client. In this case, the proxy server can further select one of methods A, B and C described above.

(16) E) If the requested data is already present in the storage device when a data acquisition request is received, the proxy server omits the step of acquiring the requested data from a server and transfers, to the client, the requested data which is already present in the storage device. In other words, the storage device is used as a cache memory.

(17) In the system of the present invention, when a data acquisition request is issued, it is sufficient if the communication channel connecting the client and the proxy server is connected only during data transfer from the server to the client [1]*, and it can be disconnected while the server is preparing the requested data and while it is transmitting the data to the proxy server. As a result, the total connection time is shorter than in a conventional system. To the extent that the connection time is shortened, the probability of disconnection in the course of data transfer is decreased and thus more stable communication can be anticipated. The proxy server can also be adapted to select a data acquisition method which is as advantageous as possible in terms of, for example, communication cost and communication time. Given these measures, the system of the present invention readily reduces communication cost.

(18) The proxy server is preferably provided with a storage device having sufficient capacity to hold all the requested data. This enables data transfer from the proxy server to the client to be carried out at a rate which is independent of the server transfer rate, and at a time that is convenient for the client. Moreover, as described above, it is feasible to use the storage device as a cache memory. It is also preferable to compress the data when transferring it from the proxy server to the client. Given these measures, further shortening of communication time and reduction of communication cost can be anticipated. It should be noted that the proxy server is typically implemented using a computer, and that a computer program for this purpose can be installed or loaded in the computer by way of a variety of means, including various kinds of disk storage, semiconductor memory, or a network.

Mode of embodying the invention

(19) FIG. 3 shows the constitution of an embodiment of the data communication system of the present invention.

(20) In this system, communication computer (i.e., proxy server) 20 having novel functions conforming to the principles of the present invention is present between client 10 for utilising data and server 30 for supplying data.

(21) Client 10 and proxy server 20 can communicate via a network 40. It is acceptable if this network 40 has narrow bandwidth communication channels, high communication cost and unstable connections. Typically, network 40 is a public switched telephone network or a mobile network, and in such cases communication channel 41 between client 10 and proxy server 20 is normally in a disconnected state and is connected whenever the need arises. On the other hand, proxy server 20 and

* Numbers in square brackets refer to Translator's Notes appended to the translation.

server 30 communicate via a separate network 50. This separate network 50 preferably has broadband communication channels, low communication cost and stable connections, and is typically a fixed network in which communication channel 51 between proxy server 20 and server 30 is always connected.

(22) In FIG. 3, only one each of client 10, proxy server 20 and server 30 is illustrated. However, typically there are a plurality of clients 10, proxy servers 20 and servers 30. Each client 10 can communicate with any of the plurality of proxy servers 20, and each proxy server 20 can communicate with any of the plurality of servers 30.

(23) Client 10 is a user-operated terminal and has application 11 and access control unit 12. Application 11 is a user-launched application program and has functions such as displaying and processing data that have been acquired from the network. Application 11 also has a function of issuing a data acquisition request in order to acquire data from the network. Access control unit 12 has a function of transferring, to proxy server 20, a data acquisition request that has been issued by application 11; a function of issuing a data transfer request to proxy server 20 after the data acquisition request has been transferred; and a function of receiving data that has been transferred from proxy server 20.

(24) The "data acquisition request" referred to above contains a data designation and a server designation (but it is not essential that it contains a server designation). For application 11, the data acquisition request is a command for causing the designated data to be acquired from the designated server (if there is no server designation, the data is acquired from a server appropriately selected by proxy server 20) and transferred to application 11. For proxy server 20 which receives this data acquisition request, the request constitutes a command to acquire the designated data from server 30 and to hold it in proxy server 20. The "data transfer request" is subsequently issued following a prior data acquisition request, and is a command for requesting that the data which proxy server 20 has acquired in accordance with the prior data acquisition request and which it is holding, is transferred to client 10.

(25) Proxy server 20 has a function whereby, in response to the data acquisition request from client 10, it acquires the designated data from designated server 30 (or, from a selected server) and stores this data. Proxy server 20 also has a function whereby, in response to the data transfer request from client 10, it transfers to client 10 the data that was acquired by means of the prior data acquisition request and that is now stored. Proxy server 20 has request acceptance unit 21, network access unit 22, storage device 23, data transfer unit 24, and interfaces 25 and 26.

(26) Request acceptance unit 21 receives a data acquisition request or a data transfer request from client 10. If it receives a data acquisition request, it interprets the received data acquisition request and passes on to network access unit 22 a data acquisition request for acquiring data from a server capable of fulfilling this request. If request acceptance unit 21 receives a data transfer request, it passes this data transfer request to data transfer unit 24.

(27) The specific operation of request acceptance unit 21 when it has received a data acquisition request is as follows. If both a data designation and a server designation are contained in the received data acquisition request, request acceptance unit 21 can pass on this data acquisition request just as it is to network access unit 22. However,

if only a data designation is contained in the received data acquisition request, and no server designation is contained therein, request acceptance unit 21 employs a decision process of its own to select an appropriate server for providing the designated data, generates a data acquisition request that contains a designation of this selected server, and passes on this request to network access unit 22. A variety of criteria can be employed for selecting an appropriate server from among several servers which are capable of providing the designated data and which are present in network 50. These criteria include selecting a server with less expensive charges; selecting a server with a higher transfer rate; selecting a server with less congested access; selecting a physically nearer server; and selecting a server where the communication channel to the server is less congested and offers a better communication state. Data relating to the state and performance of each server and to the state of communication channels, this data being fundamental for server selection, can be acquired from time to time from network 50 and held in advance in request acceptance unit 21. Alternatively, it can be acquired by interrogating individual servers when a server is to be selected. Even when a server designation is contained in a data acquisition request from client 10, if it is not favourable to use the designated server — for example, because the communication state has deteriorated due to this designated server being busy or down, or due to the communication channel to the designated server being very congested — request acceptance unit 21 can automatically select another suitable server, in similar manner to when a server designation is not contained in the data acquisition request.

(28) Network access unit 22, in response to the data acquisition request from request acceptance unit 21, acquires the designated data from designated server 30 and stores this data in storage device 23. Storage device 23 holds this acquired data.

(29) Data transfer unit 24, in response to a data transfer request from request acceptance unit 21, reads, from storage device 23, the data that has been acquired by the prior data acquisition request, and transfers this data to client 10. Interface 25 is a communication interface between proxy server 20 and client 10. Interface 26 is a communication interface between proxy server 20 and server 30.

(30) FIG. 4 and FIG. 5 are flowcharts showing, in respect of the system depicted in FIG. 3, the procedure whereby application 11 on client 10 acquires data that is present in server 30.

(31) This procedure comprises two stages. In the first stage, as shown in FIG. 4, client 10 issues a data acquisition request. In the second stage, as shown in FIG. 5, client 10 actually receives the data. These two stages can be performed consecutively within a single period of time during which communication channel 41 between client 10 and proxy server 20 is connected. However, the first stage and the second stage are preferably performed at different connection occasions as indicated by the separate flowcharts of FIG. 4 and FIG. 5, whereby the advantageousness of the present invention — namely, the decrease in communication cost — can be made still more effective.

(32) The first stage, shown in FIG. 4, will now be described.

(33) When a data acquisition request is issued from application 11, this request is passed to access control unit 12 (S11). When access control unit 12 receives this

request, it connects communication channel 41 to proxy server 20 and establishes communication (S12). Next, the data acquisition request is transferred from access control unit 12 to request acceptance unit 21 via interface 25 (S13). When the transfer of the data acquisition request from access control unit 12 to request acceptance unit 21 is completed, access control unit 12 temporarily disconnects communication channel 41 (S14).

(34) After this, in proxy server 20, when request acceptance unit 21 has interpreted the data acquisition request that has been transferred and has ascertained the designated data and (if a server designation is included in the request) the designated server, it firstly checks whether or not this designated data is already present in storage device 23 (S15). If the designated data is in storage device 23, request acceptance unit 21 ends the processing of the data acquisition request, and proxy server 20 returns to its original standby condition.

(35) On the other hand, if the designated data is not in storage device 23, request acceptance unit 21 sends the data acquisition request to network access unit 22 (S16). In this case, if a server designation was not included in the data acquisition request from client 10, or, if a server designation was included but there is a problem in accessing the designated server, then as explained above, request acceptance unit 21 selects, on the basis of the above-mentioned fundamental information acquired from network 50, an appropriate server that is capable of supplying the designated data, causes a designation for this selected server to be included in the data acquisition request, and sends this data acquisition request to network access unit 22.

(36) In accordance with the data acquisition request, network access unit 22 acquires, via interface 26, the designated data from server 30 which this request has designated (S17), and stores all this acquired data in storage device 23 (S18). This completes the first stage, whereupon proxy server 20 returns to its original standby condition.

(37) After this, the second stage depicted in FIG. 5 is performed.

(38) Firstly, access control unit 12 of client 10 reconnects communication channel 41 to proxy server 20 and establishes communication (S21). The operations involved in this reconnection are performed automatically by access control unit 12 after the elapse of a suitable time interval from the issuing of the data acquisition request. However, another feasible method is to perform reconnection at a time desired by the user, by means of an instruction from application 11. Alternatively, it is also feasible to call client 10 from proxy server 20 when data acquisition by proxy server 20 has been completed, and to establish communication on this basis. Thus a variety of methods can be adopted as regards how and at what juncture to perform the reconnection.

(39) Next, access control unit 12 issues a data transfer request. This data transfer request is transferred from request acceptance unit 21 of proxy server 20 to data transfer unit 24 (S22). In response to this, data transfer unit 24 reads the designated data from storage device 23 (S23) and transfers this data via interface 25 to access control unit 12 (S24). When the data transfer is completed, access control unit 12 disconnects communication channel 41 to proxy server 20 (S25). Next, access control unit 12 transfers the received data to application 11 (S26).

(40) The foregoing describes the series of procedures performed when client 10 acquires data from server 30. Given these procedures, the communication channel to the proxy server only has to be connected while client 10 issues a request and while it receives data. Communication channel 41 can be disconnected during the period that data are being prepared at server 30. As a result, the total communication time is shorter then hitherto, and to this extent there is less likelihood of disconnection during communication due to a deterioration in the communication conditions. Moreover, because the issuing of a data transfer request and the actual transfer of data can be separate connection occasions, if disconnection occurs in the course of data transfer, only this data transfer has to be repeated, and it is not necessary to repeat the processing again from practically the very beginning as has hitherto been the case. The foregoing features mean that communication cost can be reduced.

(41) In the system described above, once the communication between proxy server 20 and server 30 is established it remains in place while server 30 is preparing and transferring data. However, because this is a restricted area within the overall system, and because the number of individual communication instances will be very large, the utilisation of stable, high-speed broadband always-on communication channel 51 [2] is easier than with individual clients 10. [3] As a result, the cost per individual access can be considerably reduced.

(42) Moreover, in the system described above, the function of "interpreting the data acquisition request" possessed by request acceptance unit 21 in proxy server 20 enables an appropriate server to be automatically selected and accessed in response to a data acquisition request which does not designate a server or which has designated a disadvantageous server, and hence communication cost can be even more effectively limited.

(43) Because proxy server 20 is also provided with a cache function whereby data whose acquisition has been requested is temporarily stored in storage device 23 so that a second acquisition request can be responded to immediately, a further reduction in processing time and in communication cost can be anticipated.

(44) The cost of communication in communication channel 41 can be still further reduced if data are compressed before being transferred between client 10 and proxy server 20.

Effect of the invention

(45) According to the present invention, it is sufficient if a communication channel remains connected only while a client issues a request and while data are being received. As a result, communication time is shortened and the possibility of disconnection due to an unstable communication condition is decreased. Moreover, because transmission of a request and transfer of data can take place on different connection occasions, disconnection during data transfer only requires re-transfer of the data and does not necessitate repeating the processing from the very beginning. Communication cost can therefore be reduced.

(46) If in addition data are compressed prior to transfer, cost can be still further reduced.

(47) Provided that novel functions in conformity with the present invention are incorporated in the proxy servers and the access control units of the clients, it is not necessary to modify existing client applications or existing servers, and hence an existing system can be utilised. This enables the present invention to be implemented inexpensively.

Brief Description of the Drawings

FIG. 1 is a block diagram of a conventional data communication system.

FIG. 2 is a flowchart depicting the procedure whereby a client acquires data on a server in a conventional system.

FIG. 3 is a block diagram of an embodiment of the data communication system of the present invention.

FIG. 4 is a flowchart showing the procedure performed when a client issues a data acquisition request in the system depicted in FIG. 3.

FIG. 5 is a flowchart showing the procedure performed when a client receives data in the system depicted in FIG. 3.

Description of referencing numerals

10□□□□□□□□□ client
11□□□□□□□□□ application
12□□□□□□□□□ access control unit
20□□□□□□□□□ proxy server
21□□□□□□□□□ request acceptance unit
22□□□□□□□□□ network access unit
23□□□□□□□□□ storage device
24□□□□□□□□□ data transfer unit
25, 26 □□□□□ interfaces
30□□□□□□□□□ server
40□□□□□□□□□ network
41□□□□□□□□□ communication channel
50□□□□□□□□□ fixed network
51□□□□□□□□□ communication channel

FIG. 1

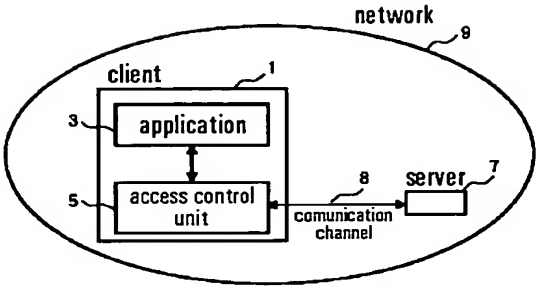

network
9
client
1
3
application
5
access control unit
8
server
7
comunication channel

FIG. 2

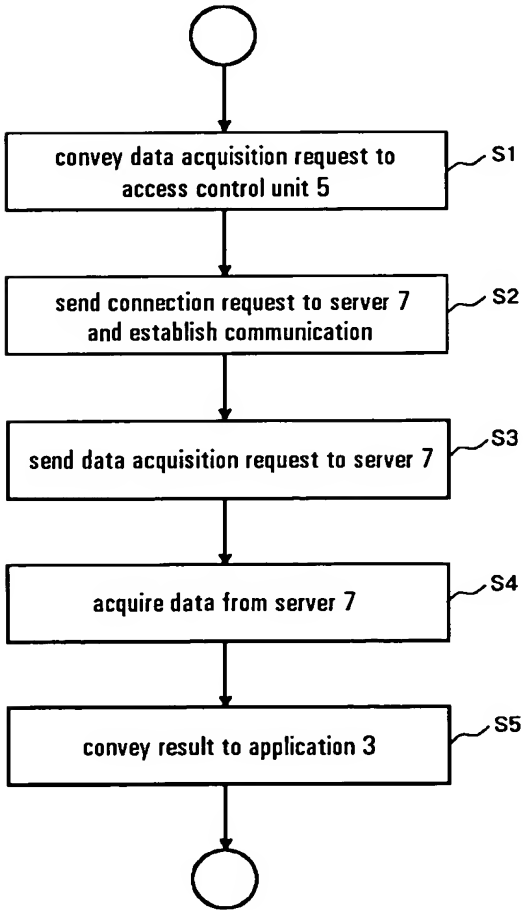

convey data acquisition request to access control unit 5
S1
send connection request to server 7 and establish communication
S2
send data acquisition request to server 7
S3
acquire data from server 7
S4
convey result to application 3
S5

FIG. 3

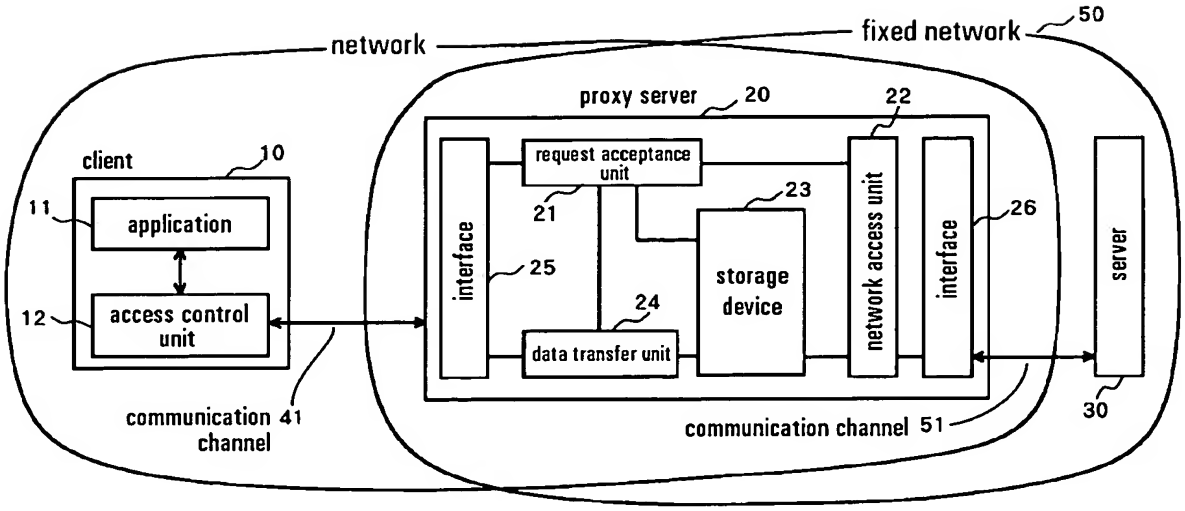

network
fixed network
50
proxy server
20
22
client
10
11
application
12
access control unit
request acceptance unit
21
23
storage device
24
data transfer unit
interface
25
network access unit
interface
26
server
30
communication channel 41
communication channel 51

FIG. 4

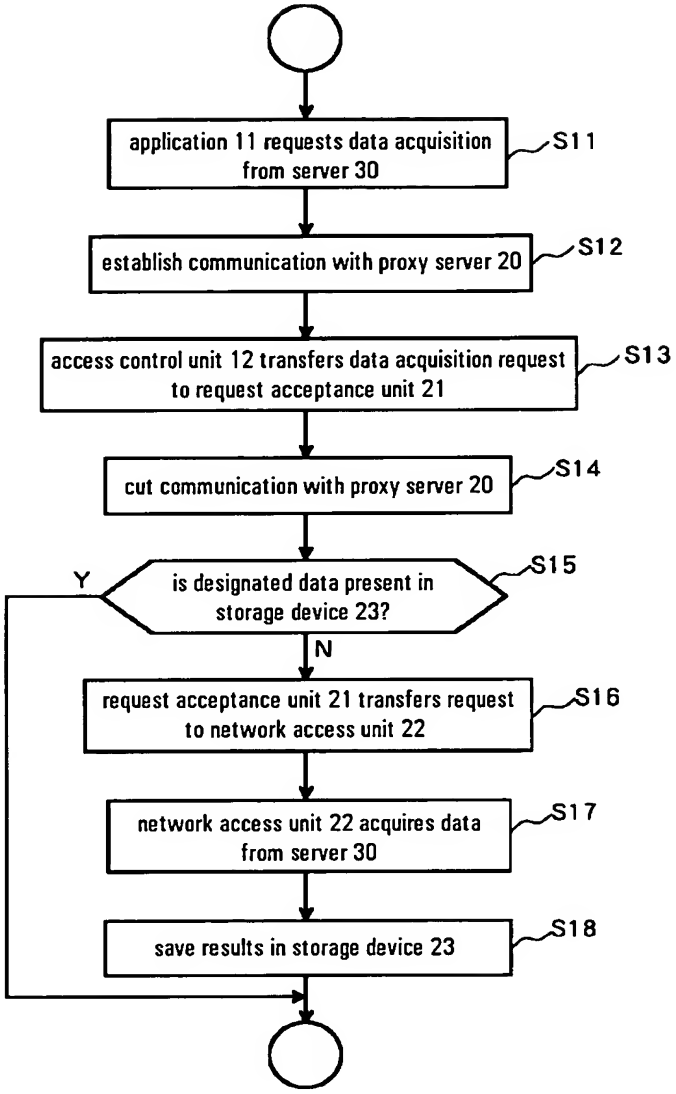
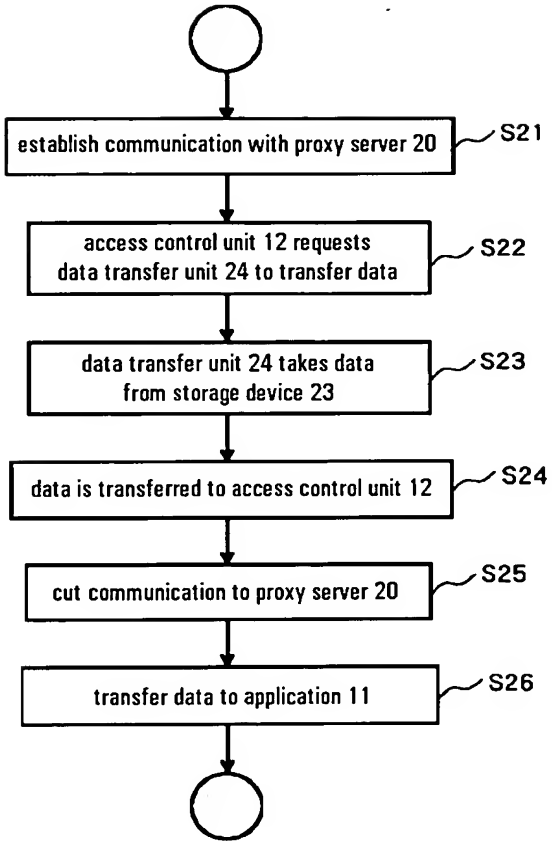

application 11 requests data acquisition from server 30
S11
establish communication with proxy server 20
S12
access control unit 12 transfers data acquisition request to request acceptance unit 21
S13
cut communication with proxy server 20
S14
is designated data present in storage device 23?
S15
Y
N
request acceptance unit 21 transfers request to network access unit 22
S16
network access unit 22 acquires data from server 30
S17
save results in storage device 23
S18

FIG. 5



TRANSLATOR'S NOTES

1. Sic. The writer presumably means "only during data transfer from the ***proxy*** server to the client".
2. The Japanese erroneously has "50" here. I have corrected this to "51".
3. Sic. It is not entirely clear what the writer means by "easier than with individual clients" in this sentence.